Irving Fisher Committee on
Central Bank Statistics

BIS

IFC-Bank of Italy Workshop on "Data science in central banking"

18-20 February 2025

Istanbul electricity demand forecast with artificial neural networks[1]

Hayriye Yasak Ozkal,
Central Bank of the Republic of Türkiye

Buse Kaylan, Meltem Sipahi and Muammer Altan Cakir,
Istanbul Technical University

[1] This contribution was prepared for the workshop. The views expressed in this publication are those of the authors and do not necessarily represent the official views of the Committee, its members, or the BIS.

Istanbul Electricity Demand Forecast with Artificial Neural Networks

Hayriye Yasak OZKAL[1], Buse KAYLAN[2], Meltem SIPAHI[3]

and Prof. Dr. Muammer Altan CAKIR[4]

Abstract

When examining the relationship between supply and demand for electricity consumption, overestimating the electrical load can lead to unnecessary energy purchases and excessive reserves, while underestimating the load can cause insufficient energy supply, increased outages, and consequently, system instability. Optimizing electricity demand forecasts to ensure efficient resource usage and reduced outages is becoming a crucial issue for energy producers, distribution companies, and consumers. This study evaluates the impact of seasonal variables such as temperature and humidity, as well as official and weekends on electricity consumption in Istanbul, using Energy Exchange Istanbul (EXIST) transparency reports. The aim is to enhance prediction accuracy by analysing the influence of these parameters on consumption forecasts. The dataset, characterized by time series features, was examined, and after applying NARX models and Prophet/LSTM hybrid models, the final model was selected based on the lowest error metrics and the highest forecast accuracy. The results indicated that Istanbul's electricity consumption is related to factors like humidity, temperature, intra-day hours, and holiday periods.

Keywords: Electricity demand forecasting, Machine learning algorithms, Prophet/LSTM model, NARX model, Generative Adversarial Networks (GAN)

JEL classification: C45, C52, L94, Q47

[1] Department of Data Governance and Statistics, Central Bank of the Republic of Türkiye,

email: hayriye.yasakozkal@tcmb.gov.tr

[2] Department of Data Engineering and Business Analytics, Istanbul Technical University,

email: kaylanb23@itu.edu.tr

[3] Department of Data Engineering and Business Analytics, Istanbul Technical University,

email: sipahigunindi23@itu.edu.tr

[4] Department of Physics Engineering, Istanbul Technical University, email: cakirmua@itu.edu.tr

1. Introduction

As the population, industry and technology-based commercial activities and the demand for technology increase in Turkiye over the years, the demand for electrical energy increases in parallel. The non-storability of electricity affects the electricity generation, distribution and consumption related planning and operational strategies.

The fact that the amount of electricity consumed is greater than the amount of electricity produced, or that the amount of electricity produced is greater than the amount consumed, leads to an energy imbalance. Accordingly, if load forecasts are overestimated, this can lead to unnecessary energy purchases and excess reserves, while if load forecasts are lower than actual consumption, this can lead to insufficient energy supply and system instability due to outages. For this reason, many decisions in the electricity markets depend on accurate forecasting of consumer demand in order to ensure the efficient operation of the energy markets, to avoid wasting resources and to reduce costs, and on accurate load forecasts that ensure a balance between supply and demand (Akman et al., 2018).

In order to ensure system efficiency in supply and demand management, electricity is traded through the Day-Ahead Market (DAM) and Intraday Market (IMM) in addition to long-term bilateral agreements. DAM tries to ensure a balance between supply and demand by planning production and consumption a day in advance. The participants submit their hourly bids and calls for the next day, and Energy Exchange Istanbul matches these offers. In cases where the energy supply-demand balance cannot be achieved through the DAM, the imbalance is attempted to be eliminated by making flexible and instantaneous transactions through the IMM. The aim is to make energy trading more efficient by maintaining the balance between supply and demand instantaneously for both producers and consumers.

In previous studies on electricity demand forecasting, perceived temperature, weekends, and holidays have been considered as parameters in electricity consumption predictions. In this study, the periodic influence of the relevant variables that affecting electricity consumption on prediction accuracy are investigated. After analysing their impact on consumption trends, a deep learning model is developed with the aim of improving prediction accuracy.

2. Materials and Methods

During the energy demand forecasting literature review to determine which model to use, it was found that long-term (year), medium-term (month), short-term (day/hour), and real-time (minute/second) forecasts can be made using historical data and appropriate forecasting models. In an examination of previous studies, it can be seen that market conditions, seasonal variables, social and demographic events are considered within the context of the research and an attempt is made to forecast the load. For this purpose, various models based on multiple linear regression analysis (MLR), artificial neural networks (ANN), ARIMA, least-squares support-vector machines (LS-SVM) methods have been developed and according to the results of the study, it is understood that ANN and LS-SVM are compatible with the actual values in the short-term forecasts (Akman et al., 2018). Therefore, NARX (Nonlinear Autoregressive External Input Model) models based on SimpleRNN (Recurrent Neural Networks) and LSTM (Long-Short Term Memory) artificial neural networks, and Prophet-LSTM hybrid models were developed using electricity consumption data for Istanbul and their short-term (1 hour) performance was tested.

2.1. Data

The data period used in this study is from January 2021 to December 2023. Hourly temperature and humidity data are obtained from the open-source web service Open-Meteo, and actual hourly electricity consumption is in megawatt hours (MWh) and is obtained from Energy Exchange Istanbul (EXIST).

2.1.1. Generating Synthetic Data

The hourly electricity consumption data obtained from EXIST is representative of Turkiye as a whole, and no province-based hourly data is available. Therefore, it is proposed to synthetically generate hourly electricity consumption data for Istanbul using Generative Adversarial Networks (GANs). The model development process consists of two stages. First, a synthetic version of Turkiye's hourly electricity consumption data was created. Subsequently, the generated data set is scaled according to the monthly consumption rates of Istanbul province. Finally, the hourly electricity consumption data of Istanbul province is obtained.

The GAN model used in first stage consists of a generator and a discriminator network. The purpose of the generator is to produce realistic looking data points from a random input vector. The Generator used in this study is configured with three dense layers of increasing units. These layers have 256, 512 and 1024 units respectively, and each layer is supplemented by the LeakyReLU activation function and batch normalisation. LeakyReLU accelerates the learning process by being particularly sensitive to negative input values, while Batch Normalisation ensures that data distributions between layers are normalised. The output layer is terminated by a single unit linear activation function. The Discriminator is designed to distinguish whether the data is real or generated by the Generator. The Discriminator is structured in three dense layers with 1024, 512 and 256 units in descending order. These layers have the LeakyReLU activation function, and Dropout (0.3) is applied to each layer to avoid overfitting. The sigmoid activation function is used in the output layer of the Discriminator. This provides an output to determine the probability of the data being real. The model is implemented using the adversarial training process, where the Generator and the Discriminator networks are trained and both iteratively attempt to outperform each other. In order to train the model long enough, this training process lasted for 10,000 epochs.

In the second stage of the study, the synthetic data generated by the GAN model is scaled according to the monthly consumption percentages of the Istanbul province in Turkiye, as provided by the Electricity Market Development Reports of the Energy Market Regulatory Authority (EMRA). The similarity between synthetic and actual patterns was evaluated, and due to observed inconsistencies, the dataset was excluded from further analysis. Instead, the hourly electricity consumption data for

Istanbul has been obtained by scaling with the province-specific consumption percentage billed throughout Turkiye, which is also given in the report.

2.2. Exploratory Data Analysis (EDA)

In order to understand the variability of energy demand over time and the impact of factors such as seasonality, holidays and working hours on demand, hourly electricity consumption data for the province of Istanbul from January 2021 to December 2023 was analysed. Density plot of this data, as shown in Figure 1, shows that the consumption values generally vary between 5,000 and 7,000 MW, but there are also some values outside this range. When the probability distribution is analysed (see Figure 2), it is evident that while the data appear to be approximately normally distributed, significant deviations are observed, particularly at the extreme points.

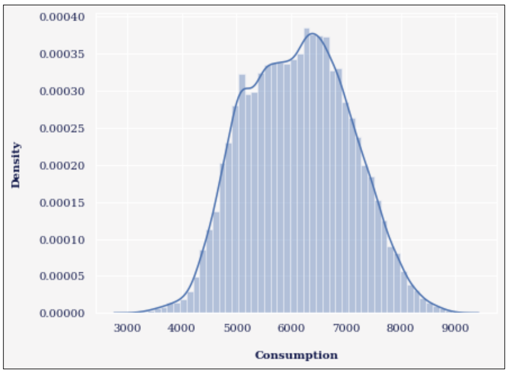


Figure 1. Density Plot

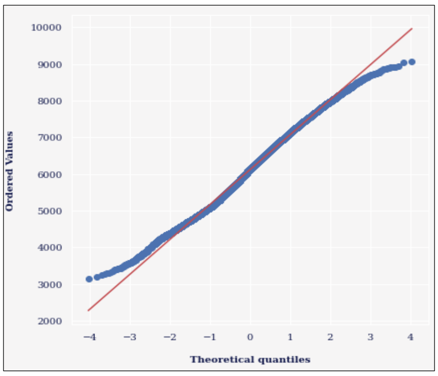


Figure 2. Probability Plot

Analysis of the Istanbul's hourly electricity consumption graph as illustrated in Figure 3, shows that there are significant peaks in electricity consumption in the middle and toward the end of the year. Given that the observed periods coincide with the summer and winter months, seasonal effects are considered to be significant. During specific periods, such as religious holidays, a notable decline in electricity consumption is observed (indicated by the red arrows). These decreases may be the result of the seasonal slowdown in economic activity that typically occurs during the holiday period.

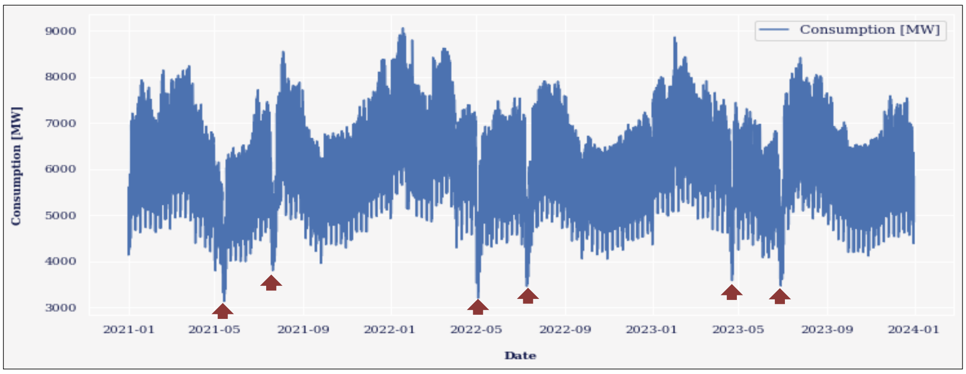


Figure 3. Istanbul Hourly Electricity Consumption Plot

An analysis of seasonal trends in electricity consumption (Figure 4) shows that there has been a significant increase in electricity consumption, particularly in the winter (December-January) and summer (July-August) months. Conversely, during the spring and autumn seasons, consumption tends to be more balanced and at lower levels. The cyclical patterns that regularly recur in the data are assessed to be explained by factors such as seasonal effects and operational calendars.

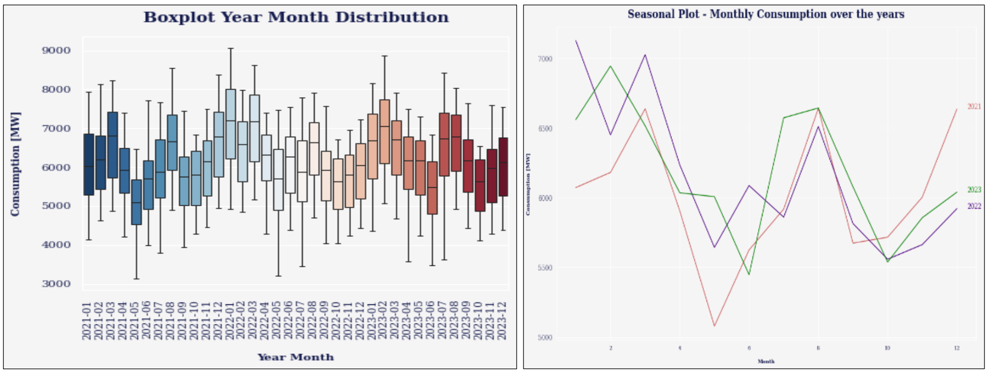


Figure 4. Istanbul Monthly Electricity Consumption Distribution Charts

As illustrated in Figure 5, electricity consumption is higher on weekdays due to the intensity of industrial and commercial activities. On the weekends, the decrease in these activities leads to a decrease in the total consumption of electricity.

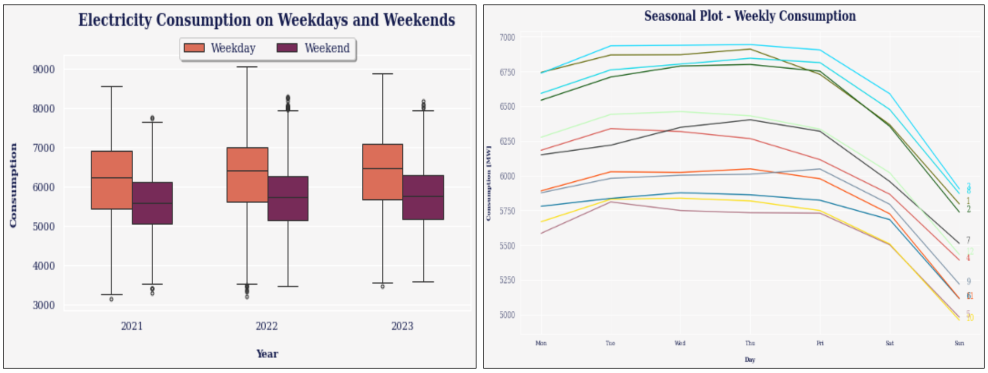


Figure 5. Istanbul Electricity Consumption Charts - Weekday/Weekend and Weekly

Analysis of the hourly electricity consumption trend reveals a sudden increase in consumption around 5:00-6:00 a.m., which continues until 11:00 a.m. This increase is followed by a gradual decrease in consumption until 17:00-18:00 p.m. (see Figure 6).

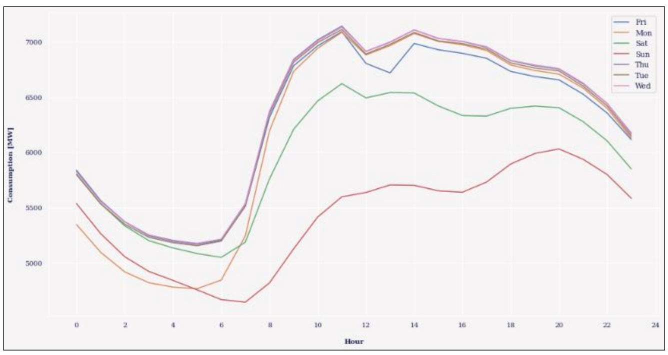


Figure 6. Istanbul Hourly Electricity Consumption Plot

Within the scope of the statistical analysis, the autocorrelation (ACF) and partial autocorrelation functions (PACF) of Istanbul hourly electricity consumption data were analysed, and it was observed that there are regular jumps in the autocorrelation graph for every 24 cycles. Through the analysis of the PACF plot, it is apparent that there are notable autocorrelations for the initial lags.

In this study, artificial neural network (ANN) models are preferred due to their superior performance in resolving complex relationships in nonlinear multivariate data sets when compared to traditional time series methods (Cheng et al., 2017). The single step (1 hour later) forecasting approaches utilized in this context are RNN and LSTM-based NARX models and the Prophet-LSTM hybrid model.

2.3.Overview of Forecasting Models

2.3.1. NARX Neural Network

The nonlinear autoregressive network with exogenous inputs (NARX), which is widely used in time series modelling, uses the past values of the dependent variable as well as the past values of the exogenous explanatory variables to predict the future value of the dependent variable. The mathematical equation used in the NARX model is as follows:

$$y(t)= f(y(t-1),y(t-2),\ldots,y(t-n_y),u(t-1),u(t-2),\ldots,u(t-n_u))$$ [5]

2.3.2. Long Short-Term Memory

LSTM (Long Short-Term Memory) is a specialised type of Recurrent Neural Network (RNN) that is widely used in time series due to its capacity to learn long-term relationships within consecutive time steps. In recurrent neural networks, inputs are received, hidden states are updated based on previous computations, and predictions are generated for each element of the sequence. The LSTM learns both short-term and long-term dependencies at each time step using a memory cell structure, unlike the RNN.

The memory cell is composed of a self-recurrent neuron and three gates: the input gate, the output gate, the forget gate. The gates are is responsible for regulating the interactions between adjacent memory cells and the memory cell itself. The input gate controls whether the input signal can change the state of the memory cell. The output gate regulates the memory cell's state based on if it can change the state of the other memory cell. In addition, the forget gate chooses whether the previous state of the cell is remembered or forgotten (Cocianu and Avramescu, 2020).

[5] The symbol y(t) represents the predicted value (model output). The variables y(t-1), y(t-2) refer to the lagged values of the dependent variable, u(t-1), u(t-2) to the lagged values of the exogenous inputs. ny is the model feedback for the dependent variable, nu is the model feedback for the exogenous inputs, and f is the learned nonlinear function.

2.3.3. Prophet Model

Facebook Prophet is an open-source library developed for the purpose of forecasting time series with complex features including trends, seasonality, and holidays. It is included in the study because it successfully handles trend shifts and outliers in time series. The mathematical equation for Prophet is expressed as follows (Bashir et al., 2022):

$$y(t)=g(t)+s(t)+h(t)+\epsilon t$$ [6]

2.4. Proposed Methodology

Based on the results of the data exploration and analysis, the final dataset was created using electricity consumption data and related factors (Table 1). The dataset covers the time interval from 00:00 on January 1, 2021 to 23:00 on December 31, 2023.

Selected Factors — Table 1
Apparent Temperature
Perceived Humidity
Holiday
Weekend
Related Electricity Consumption Hour
Month of the year

The final dataset for NARX model development is used to create two separate series, the target variable (y) and the attributes (X) then series divided into three parts: 60% training data, 20% validation data for tuning model hyperparameters, and 20% test data for evaluating model performance. The chronological order has been taken into consideration when division is performed. All recorded time series values were then transformed to a normal distribution using the standard MinMaxScaler. The lags of the normalized series up to 24-time steps determined by the PACF plot are added to the feature series and formatted according to the configuration of each model's input layer.

The NARX RNN model used in the study is composed of three layers: an input layer, a SimpleRNN layer with 50 units and a ReLU activation function, and an output layer with one unit for the estimation of the electricity consumption. The model was compiled using the Adam optimizer to adjust the learning rate as the data was trained, and the Mean Squared Error (MSE) loss function to minimize the mean squared differences between predicted and actual values. In order to preserve the

[6] The variable y(t) is the quantity that is being estimated, g(t) is the trend function responsible for modelling non-periodic variations, s(t) is the seasonality (hourly, daily, weekly, yearly), h(t) is the holiday/special day effect, and ϵt is the error term, i.e., the random noise or error that cannot be modelled.

chronological sequence of data, the training of the model was performed in batches of 32 instances for 50 cycles without shuffling and to develop an efficient model, the Early Stopping method was used (Table 2).

Differently from the NARX RNN model, the LSTM-based NARX model has two sequential LSTM layers with 50 units and a dropout layer between them, which disables a random 20% of neurons during training to prevent overfitting.

Summary of Models

Table 2

Model Name	Layer (type)	Output Shape	Parameters	Total Parameters
NARX RNN Model	simple_rnn (SimpleRNN)	(None, 50)	2,900	2,951
	dense (Dense)	(None, 1)	51	
LSTM Based NARX Model	lstm (LSTM)	(None, 24, 50)	11,600	31,851
	dropout (Dropout)	(None, 24, 50)	0	
	lstm_1 (LSTM)	(None, 50)	20,200	
	dense (Dense)	(None, 1)	51	
Prophet LSTM Hybrid Model	Prophet Model	(26.256, 32)	0	31,851
	lstm (LSTM)	(None, 24, 50)	11,600	
	dropout (Dropout)	(None, 24, 50)	0	
	lstm_1 (LSTM)	(None, 50)	20,200	
	dense (Dense)	(None, 1)	51	

The objective of developing the Prophet-LSTM hybrid model is to use the Prophet model as the main prediction tool, and feed the remaining errors from the predictions into the LSTM model for more accurate estimation. This method allows LSTM to learn more complex and short-term patterns that Prophet cannot model.

In hybrid model approach, the data set has initially prepared in the input format of the Prophet model, with the time label (ds) containing the relevant time and date of the consumption, the hourly electricity consumption data as the target variable (y), and the lags of the target variables up to 24-time steps and the independent variables designated as exogenous inputs. Exogenous inputs are added to the Prophet model as regressors, and initial consumption predictions are obtained using seasonal and trend components. In the next step, the residuals are calculated based on the actual and predicted consumption values. Subsequently, pre-processing and modelling steps similar to the LSTM-based NARX model are applied to residuals and exogenous inputs, except that ReLU is used instead of the LSTM's default activation function. Thus, the residuals are estimated using delayed values and exogenous inputs.

Lastly, the residuals estimated by the LSTM stack and the trend (*T(t)*), seasonality (*S(t)*) and other components (extra_regressors_additive) estimated by Prophet are added to obtain the final electricity consumption forecast.

$$y(t) = T(t) + S(t) + R(t) + beta * regressor(t)$$

3. Experimental Results

The statistical performance of the NARX RNN, LSTM-based NARX, and Prophet-LSTM hybrid models developed for the forecasting Istanbul's hourly electricity consumption data is measured by the root mean square error (RMSE) and the mean absolute error (MAE) error metrics (Table 3).

Performance Evaluation of the Models — Table 3

MODELS	PERFORMANCE METRICS			
	Train		Test	
	MAE	RMSE	MAE	RMSE
NARX RNN Model	86.65	115.68	90.39	116.87
LSTM Based NARX Model	133.41	168.11	117.48	148.66
Prophet Model	82.93	111.95	85.88	112.22
Prophet-LSTM Hybrid Model	62.17	80.48	65.58	84.73

Evaluating the performance of the models over training and test data shows that the models can generalize the data, except for the LSTM-based NARX model. According to the performance results, the Prophet-LSTM hybrid model outperformed the other models by achieving the lowest error rates on both training and test data.

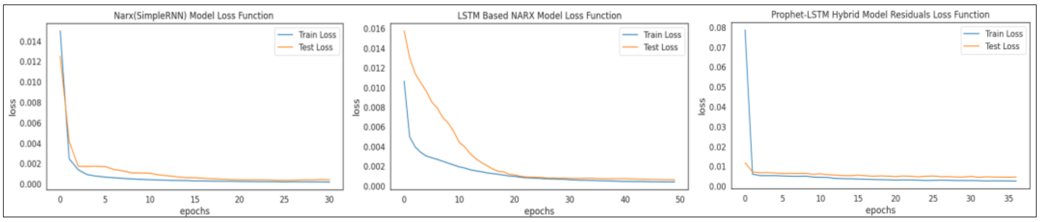


Figure 7. Comparison of Models' Loss Function Graphs

Comparing the loss functions of the models used in the study, it is observed that the training and test losses of all models decreased rapidly in the first cycles but remained stable after a while (Figure 7). As both training and test losses stabilized at similar levels, the potential for overfitting or underfitting was not evident in any of the models.

4. Conclusions and Suggestions for Further Work

Considering the changing social expectations, increasing electricity consumption needs in technology and industry, it is expected that energy imbalance problem will continue in the near future and accordingly, optimization studies in forecasting electricity demand will be crucial. In this study, the optimization problem is investigated through the use of artificial neural networks and hybrid models within the context of the factors that affect the consumption for the province of Istanbul.

According to the study, evaluating the performance metrics and loss function plots together, the Prophet-LSTM hybrid model is more successful in predicting hourly electricity consumption compared to the other models. Also, the high error metric values in the training data indicate that the LSTM-based NARX model could not fully adapt to the current data set. One of the findings of the study is that synthetic data sets may not replicate time series which contain seasonality, trends, and are affected by exogenous factors as in our case.

We conclude that the results are encouraging and that future efforts should be made to evaluate short- and long-term forecasting performance, improve the model performance by selecting shorter and more recent periods, periodic retraining the model with updated consumption patterns, and recalibrating exogenous variables for adapting model to changing circumstances.

References

Akman, T., Yılmaz, C., & Sönmez, Y. (2018). Elektrik yükü tahmin yöntemlerinin analizi [Analysing electricity load forecasting methods]. *Gazi Journal of Engineering Sciences, 4*(3), 168-175. [in Turkish]

Bashir, T., Chen, H., Tahir, M., & Zhu, L. (2022). Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN. *Energy Reports*, 8, 1678–1686.

Biçen, Y. (2016). Türkiye elektrik enerjisi piyasası gelişim süreci: Gün öncesi ve dengeleme güç piyasası özellikleri [Development process of the Turkish electricity market: features of day-ahead and balancing power markets]. *Karaelmas Fen Ve Mühendislik Dergisi*, *6*(2), 432-438. [in Turkish]

Cheng, Y., Xu, C., Mashima, D., Thing, V.L.L., & Wu, Y. (2017). PowerLSTM: Power demand forecasting using long short-term memory neural network. In *2017 International Conference on Advanced Data Mining and Applications. (Springer)*. (pp. 727–740).

Cocianu, C. L., & Avramescu, M.-S. (2020). The use of LSTM neural networks to implement the NARX model: A case study of EUR-USD exchange rates. *Informatica Economica*, *24*(1), 5–14.

Demirtaş, I., & Tarı, EI. (2021). Türkiye'de elektrik piyasasının rekabete açılması ve rekabetçi elektrik piyasasında elektrik tüketiminin belirleyicileri: ARDL sınır testi [Liberalization of the electricity market in Turkiye and determinants of electricity consumption in a competitive market: ARDL bounds test]. *Journal of Emerging Economies and Policy* *6*(1), 148-164. [in Turkish]

Gündüz, S., Uğurlu, U., & Öksüz, İ. (2021). Gün içi piyasası elektrik fiyat tahmini için eksik verilerin tamamlanması [Completion of missing data for intraday market electricity price forecasting]. *Avrupa Bilim Ve Teknoloji Dergisi(25)*, 334-340. [in Turkish]

Karaçor, Z., & Güvenek, B. (2010). Enerji piyasası reformlarının elektrik enerjisi piyasasına etkisi: EÜAŞ ve ayrıcalıklı şirketler üzerine bir analizi [The impact of energy market reforms on the electricity market: an analysis on EÜAŞ and privileged companies]. *Yönetim Ve Ekonomi Dergisi, 17*(1), 147-166. [in Turkish]

Kaysal, K., Akarslan, E., & Hocaoğlu, F. O. (2022). Türkiye kısa dönem elektrik yük talep tahmininde makine öğrenmesi yöntemlerinin karşılaştırılması [Comparison of machine learning methods in Turkiye short-term electricity load forecasting]. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi, 9*(2), 693-702. [in Turkish]

Kızıl Voyvoda, Ö., & Voyvoda, E. (2019). Türkiye'de enerji sektörünün yeniden yapılandırılması sürecinde hukuk düzenlemeleri – elektrik sektörü [Legal regulations in the restructuring process of the energy sector in Turkiye - electric sector]. *Çalışma Ve Toplum, 1*(60), 127-154. [in Turkish]

Massaoudi, M., Chihi, I., Sidhom, L., Trabelsi, M., Refaat, S.S., & Oueslati, F. (2019). A novel approach based deep RNN using hybrid NARX-LSTM model for solar power forecasting [Preprint]. *arXiv*. https://arxiv.org/abs/1910.10064

Yıldız, D., Yıldız, G., & Demirci, S. (2023). Fındık üretimi Için öğrenme tabanlı verim tahmini [Learning based yield forecasting for hazelnut production]. *Black Sea Journal of Engineering and Science, 6*(2), 117-126. [in Turkish]

Zuazo, I.Z., Boussaada, Z., Aginako, N., Curea, O., Camblong, H., & Sierra, B. (2021, September 8). Short-Term load forecasting of building electricity consumption using NARX neural networks model. In *2021 6th International Conference on Smart and Sustainable Technologies (SpliTech)*. (pp. 1–6).

Enerji Piyasası Düzenleme Kurumu. (2021-2023). Elektrik piyasası aylık sektör raporları [Monthly electricity market sector reports]. Retrieved April 25, 2025, from https://www.epdk.gov.tr/detay/icerik/3-0-23/elektrikaylik-sektor-raporlar [in Turkish]

ISTANBUL ELECTRICITY DEMAND FORECAST WITH ARTIFICIAL NEURAL NETWORKS

Hayriye Yasak OZKAL, Buse KAYLAN, Meltem SIPAHI
Advisor: Prof. Dr. Altan ÇAKIR

19.02.2025

Rome, Italy

Disclaimer

The views are those of the authors and do not necessarily reflect the views of the Central Bank of the Republic of Türkiye or its board members

Content

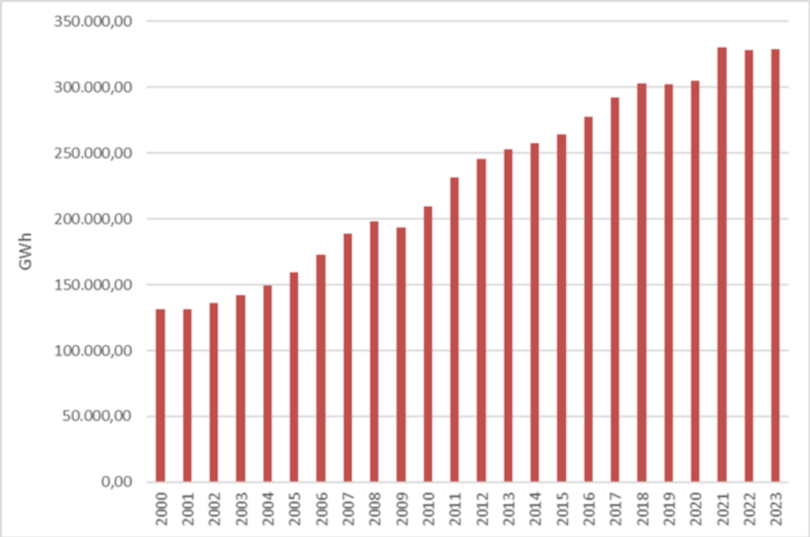
- **Introduction**
- **Main Objectives**
- **Methodology**
  - Data Collection
  - Generating Synthetic Data
  - Exploratory Data Analysis (EDA)
  - Time Series and Variable Selection
  - Data Preprocessing
  - Model Development
  - Model Evaluation
- **Key Findings & Future Work**

Introduction

Number of Consumers

Years	2021	2022	2023	2022-2023 Change
Istanbul	8.346.307	8.482.702	8.594.757	1.32 (%)
Türkiye	47.311.986	48.563.459	49.726.481	2,39 (%)

Development of Actual Consumption by Years (GWh)



As examined in the Energy Market Regulatory Authority (EMRA) 2023 sectoral report, the demand for electricity consumption and the number of consumers have increased over the years, ignoring minor deviations.

TÜRKİYE CUMHURİYET MERKEZ BANKASI

Main Objectives

- Ensuring System Stability
- Preventing Energy Shortages
- Minimizing Cost Increases
- Reducing Environmental Impact
- Preventing Resource Waste

The main objective of energy demand instability projects is to understand, manage and solve the problems caused by inaccurate demand forecasts and the resulting imbalances.

Methodology

Data Collection

- Gathering hourly humidity and temperature data for Istanbul
- Gathering hourly electricity consumption data for Turkey from January 2021 to December 2023
- Generation of Synthetic Data for Istanbul Electricity Consumption (Failed)
- Ratio of Istanbul's monthly electricity consumption to Turkey.

Exploratory Data Analysis (EDA)

- Identify patterns, trends, and correlations between electricity consumption and external factors.
- Examine historical data to understand consumption variations over time.

Model Development

- NARX RNN MODEL
- NARX LSTM MODEL
- PROPHET – LSTM HYBRID MODEL

Performance Evaluation

- Performance Metrics (MAE, RMSE)
- Model Loss Function
- Comparison of Actual and Estimated Values

Methodology

Data Collection

EPİAŞ

Actual Electricity Consumption

Temperature Humidity

OpenMeteo

In addition to these data collected on an hourly basis, information such as religious and public holidays, weekends and working days have also been added to the data set.

Methodology

Generating Synthetic Data For Istanbul Electricity Consumption

Data Source

- Total electricity consumption data of Turkey (2021-2023) from Energy Exchange Istanbul (EPIAS)
- Data includes date, hour, and consumption amount in MWh
- Monthy consumption percentages for Istanbul were used to scale the generated data

Challenge

- Lack of city-specific electricity consumption data
- Need for synthetic data to represent Istanbul's consumption

Methodology

GAN Model Architecture

GENERATOR

- **Dense Layer (256 units)** LeakyReLU Activation
- Batch Normalization
- **Dense Layer (512 units)** LeakyReLU Activation
- Batch Normalization
- **Dense Layer (1024 units)** LeakyReLU Activation
- Batch Normalization
- **Output Layer: Dense Layer (1 unit)** Linear Activation

DISCRIMINATOR

- **Dense Layer (1024 units)** LeakyReLU Activation Dropout (0.3)
- **Dense Layer (512 units)** LeakyReLU Activation Dropout (0.3)
- **Dense Layer (256 units)** LeakyReLU Activation Dropout (0.3)
- **Output Layer: Dense Layer (1 unit)** Sigmoid Activation

TRAINING PROCESS

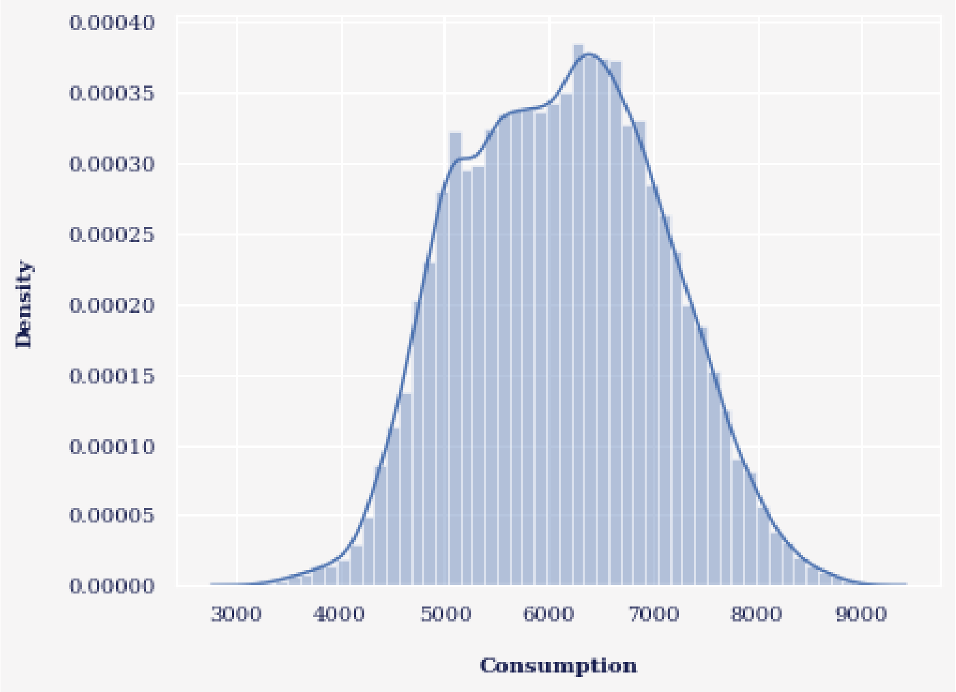
- **Epochs:** 10,000
- **Adversarial Training:** Generator and Discriminator iteratively trained to improve quality of synthetic data.

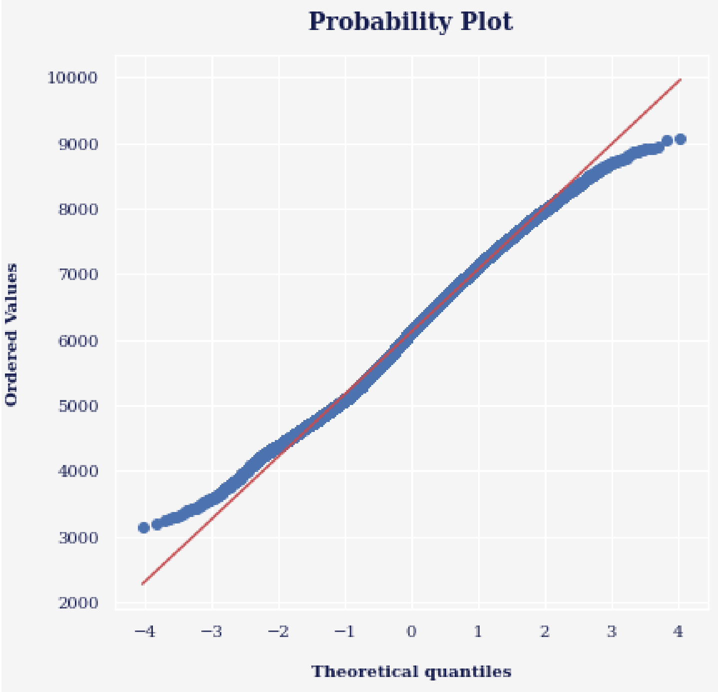
When the synthetic data generated by the GAN model is compared with the real electricity consumption data, it is observed that the synthetic data cannot simulate the real data pattern.

Methodology

Exploratory Data Analysis (EDA)

- The hourly consumption data of Istanbul province used in the analyses and modelling were obtained by scaling with the province-based consumption percentages in Turkey in the monthly reports of EMRA.
- Electricity consumption is usually concentrated between 5000-7000 MWh and extremely low/high consumption is rare.
- The data are approximately normally distributed but deviations are observed especially at the endpoints.



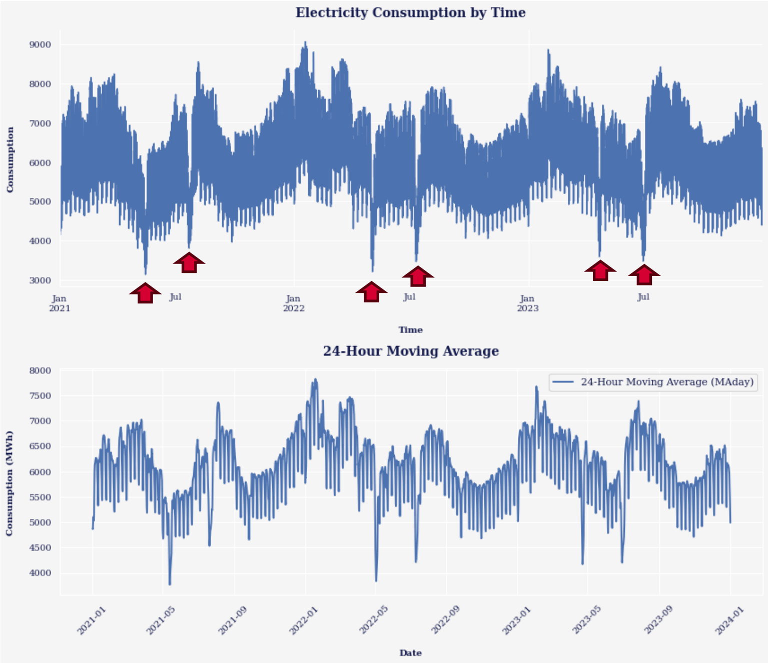


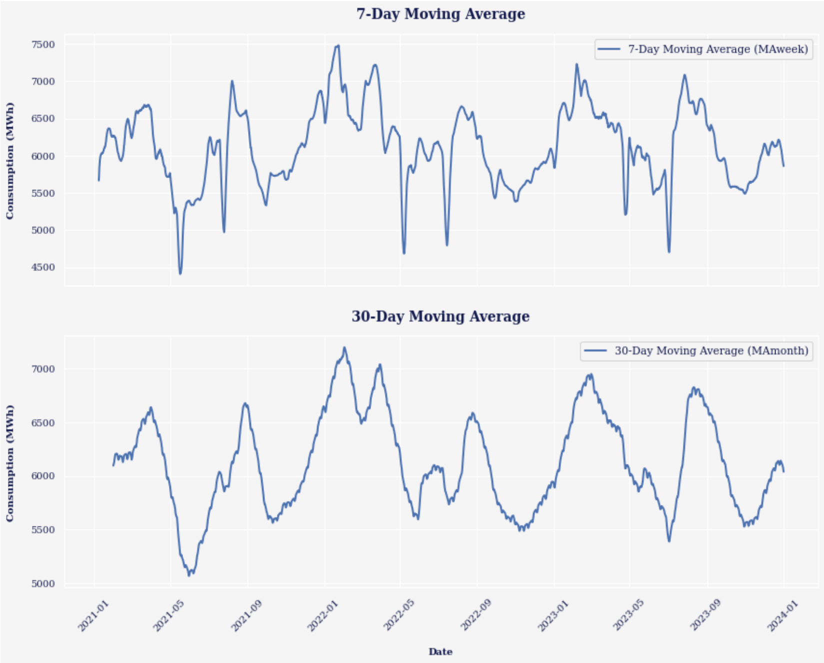
TÜRKİYE CUMHURİYET MERKEZ BANKASI

Methodology

Exploratory Data Analysis (EDA)

- When the hourly electricity consumption data of Istanbul between January 2021 and December 2023 are analysed, it is seen that consumption increases in winter months and there are fluctuations in summer months.
- It is seen that significant decreases occur in the data during religious holidays.
- The 30-Day Moving Average graph shows that the increases and decreases in overall energy consumption are seasonally significant





TÜRKİYE CUMHURİYET MERKEZ BANKASI

Exploratory Data Analysis (EDA)

Monthly Consumption

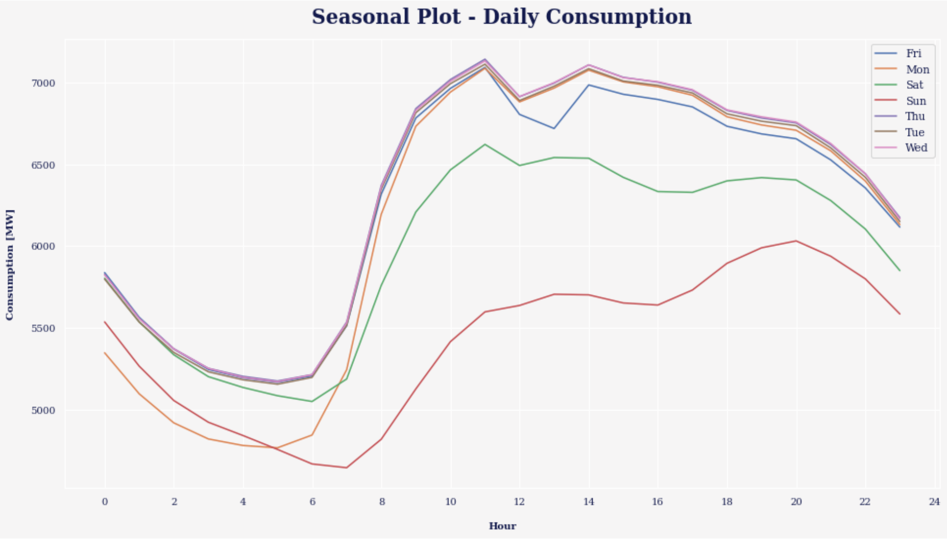
- When the monthly electricity consumption graph is analyzed, it is observed that consumption increases in summer and winter months and decreases in spring months.
- 2021 shows a different trend in terms of consumption, which may be due to the pandemic and the mild winter compared to other years

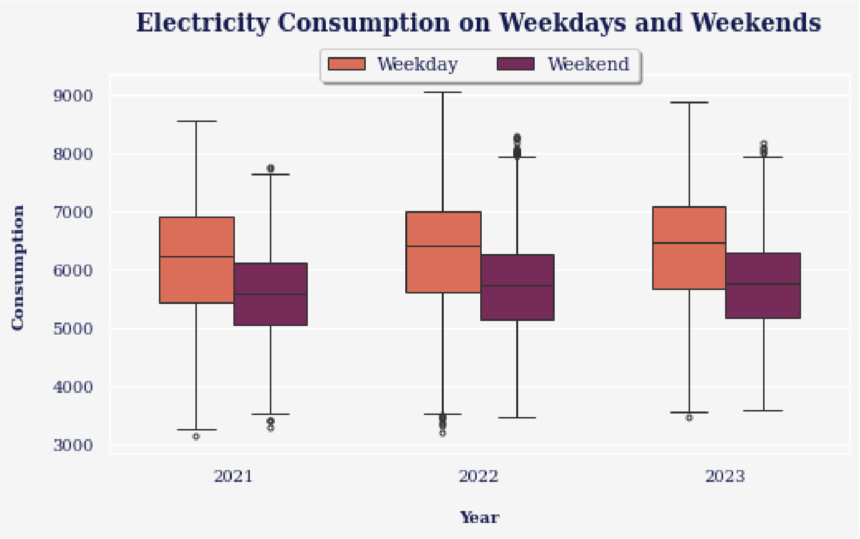
TÜRKİYE CUMHURİYET MERKEZ BANKASI

Exploratory Data Analysis (EDA)

Hourly & Daily Consumption

- When the electricity consumption graph is analysed in terms of days of the week there is a clear distinction between weekday and weekend consumption.
- For all days, there is a sharp increase in consumption starting around 5-6 AM, peaking nearly 11 AM.
- There is a gradual decline in consumption starting from late afternoon, around 5-6 PM, continuing through the evening and night.





TÜRKİYE CUMHURİYET MERKEZ BANKASI

Methodology

Time Series Analysis &

Variable Selection

- **Stationary:** The stationarity of the independent variables to be used in the model was measured by Augmented Dickey Fuller, Phillips-Perron and KPSS tests and those that failed were eliminated.
- **Correlation Matrix:** The correlation values of the dependent variable and all independent variables are observed and the independent variables to be eliminated are decided with the help of the threshold value in the decision mechanism.
- **Multicollinearity:** The VIF value used to examine whether there is multicollinearity between the explanatory variables in the data set.
- **PACF (Lag Analysis):** The appropriate number of lags to include in autoregressive (AR) models are determined by this analysis.

Selected Factors
Sensed Temperature
Perceived Humidity
Holiday
Weekend
Related Electricity Consumption Hour
Month of the year

Methodology

Data Preprocessing

- Based on the results of data exploration and analysis, a final data sets were created for electricity consumption data and related factors, with the time interval from 00:00 on January 1, 2021 to 23:00 on December 31.
- The dataset is divided into 60% training data, 20% validation data to tune model hypermeters and 20% test data to measure the model performance in chronological order.
- Next, all the time series recorded values are normalized using standart MinMaxScaler.
- Lags of the normalized series up to 24 time steps determined by PACF were added to the feature series.

Methodology

Model Development

Narx RNN Model : The Nonlinear AutoRegressive model with exogenous inputs (NARX) is a powerful modeling technique for time series forecasting.

Narx LSTM: LSTM (Long Short Term Memory) is a specialized type of RNN (Recurrent Neural Network) widely used in time series due to its capacity to learn long-term relationships within consecutive time steps.

Prophet - LSTM Hybrid Model : Combines the strengths of two popular time series forecasting techniques. The idea behind this hybrid approach is to leverage the advantages of both methods to improve the accuracy and robustness of time series forecasts.

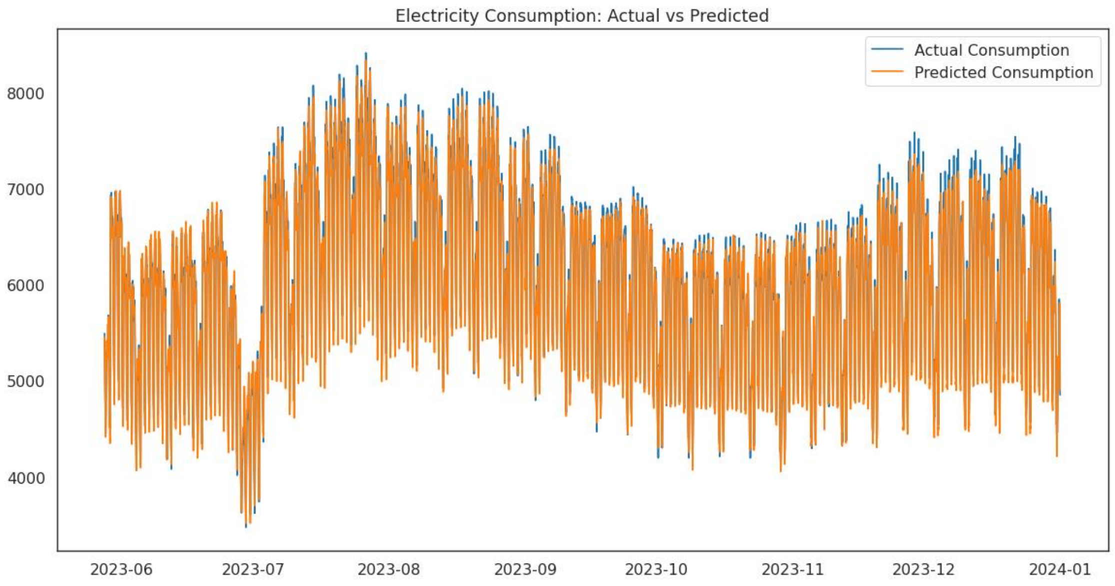
Prophet: Prophet is an open source library for forecasting time series with complex features such as trends, seasonality and holidays.

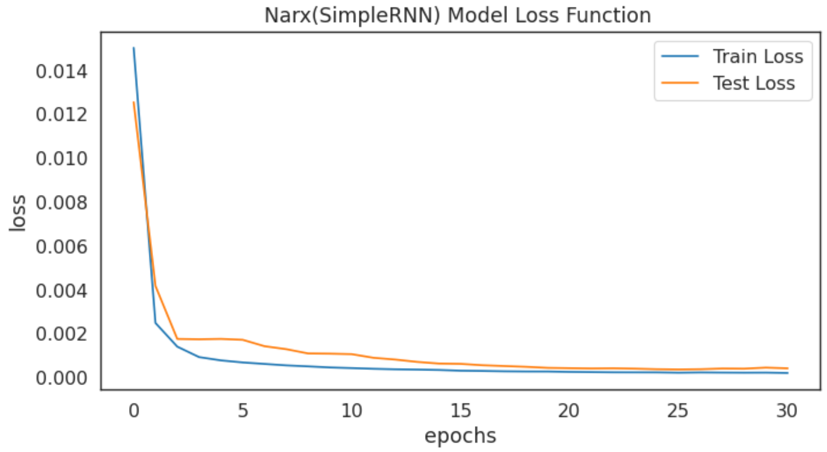
LSTM: Can learn more complex and short-term patterns that Prophet cannot model.

Model Name	Layer (type)	Output Shape	Parameters	Total Parameters
NARX RNN Model	simple_rnn (SimpleRNN)	(None, 50)	2,900	2,951
	dense (Dense)	(None, 1)	51	
LSTM Based NARX Model	lstm (LSTM)	(None, 24, 50)	11,600	31,851
	dropout (Dropout)	(None, 24, 50)	0	
	lstm_1 (LSTM)	(None, 50)	20,200	
	dense (Dense)	(None, 1)	51	
Prophet LSTM Hybrid Model	Prophet Model	(26.256, 32)	0	31,851
	lstm (LSTM)	(None, 24, 50)	11,600	
	lstm_1 (LSTM)	(None, 50)	20,200	
	dense (Dense)	(None, 1)	51	

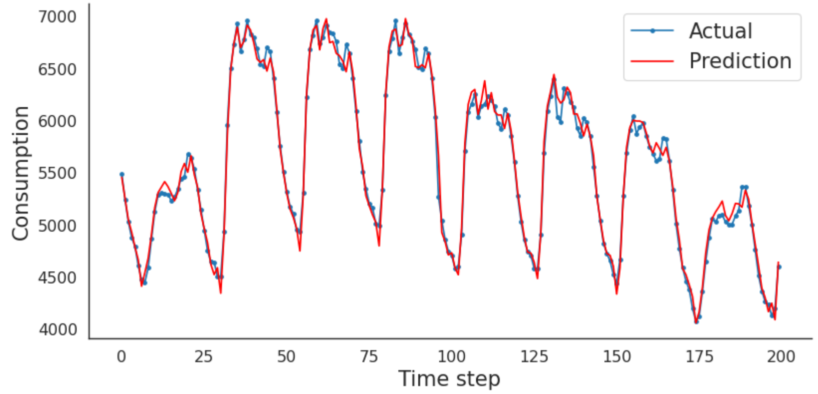
Model Development

NARX RNN Model





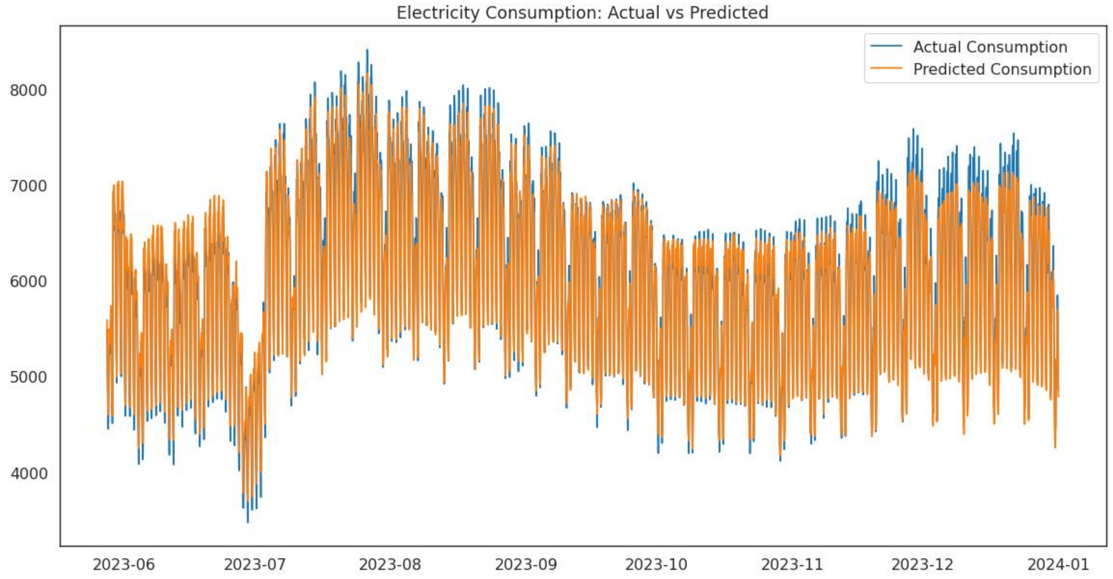
- Lag: 24
- Model: SimpleRNN (50 units, ReLU)

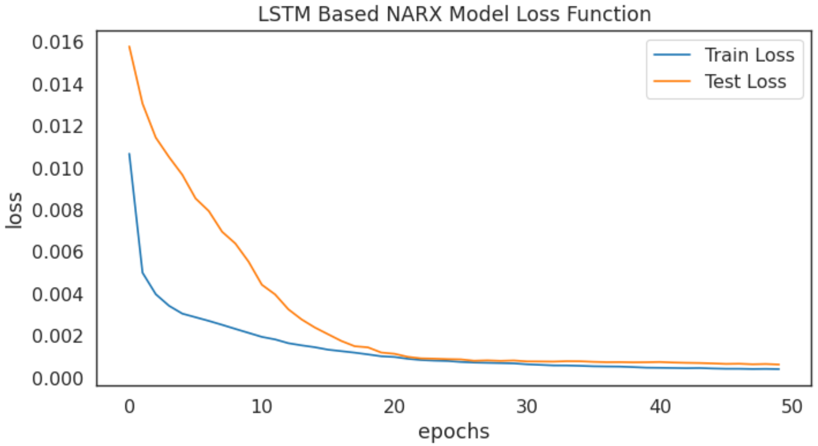


TÜRKİYE CUMHURİYET MERKEZ BANKASI

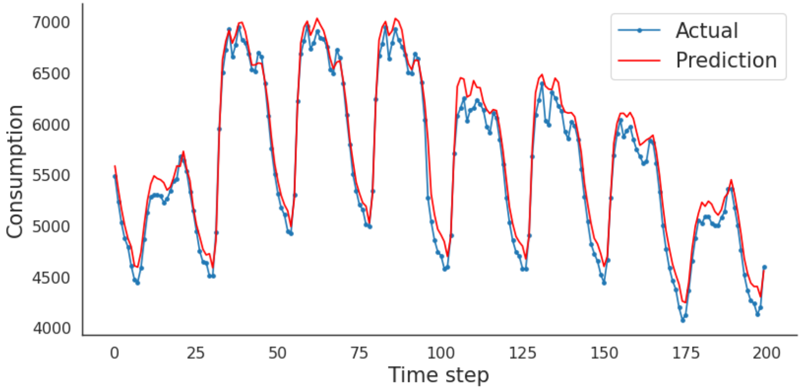
Model Development

LSTM Based NARX Model



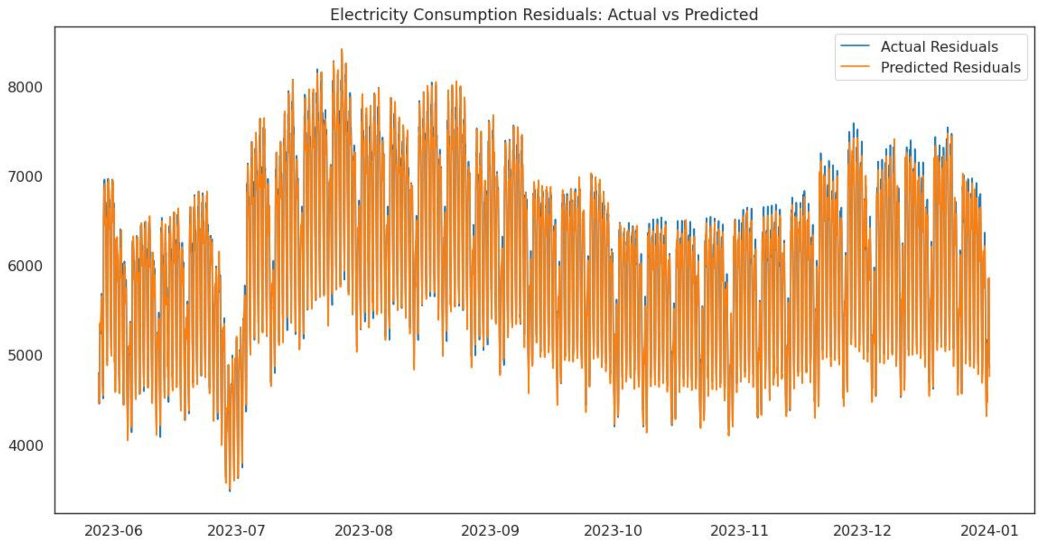


- Lag: 24
- Model: LSTM(50 units)

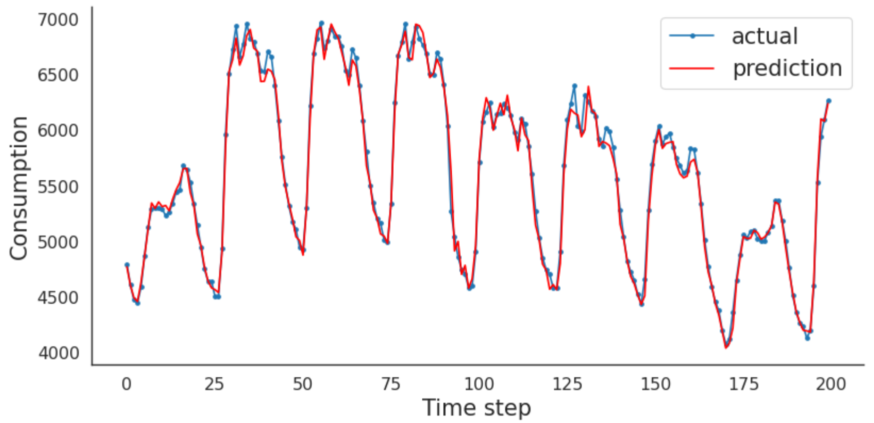


Model Development

Prophet-LSTM Hybrid Model



[21]:
```python
predicted_values=pd.DataFrame()

predicted_values[['ds', 'trend', 'daily', 'weekly', 'yearly', 'extra_regressors_additive']] = prophet_predictions[['ds', 'trend', 'daily', 'weekly', 'y

y_pred_flatten = np.full(len(predicted_values), np.nan)

# y_pred değerlerini prophet_predictions ile hizalayarak ekleyin
y_pred_flatten[-len(y_pred):] = y_pred.flatten()

y_pred_flatten[-len(y_pred):] = y_pred.flatten()

y_pred_flatten[:-len(y_pred)] = y_pred_train.flatten()

predicted_values['residual'] = y_pred_flatten

# Sezonluk ve diğer bileşenleri toplayarak y_pred sütununu oluşturma
predicted_values['y_pred'] = (
    predicted_values['trend'] +
    predicted_values['daily'] +
    predicted_values['weekly'] +
    predicted_values['yearly'] +
    predicted_values['extra_regressors_additive'] +
    predicted_values['residual']
)
```



- **Model:** Using Prophet initial predictions generated and these predictions used as input to train an LSTM model. The LSTM model learns from the residuals (the differences between the actual values and the Prophet predictions) and generates refined forecasts.

TÜRKİYE CUMHURİYET MERKEZ BANKASI

Model Development

Model Evaluation

- MAE: Measures the mean of error magnitudes, ignores direction, is easy to interpret and has low sensitivity to outliers.
- RMSE: Expresses errors in original data units, gives more weight to large errors and is more sensitive to outliers.

MODELS	PERFORMANCE METRICS			
	Train		Test	
	MAE	RMSE	MAE	RMSE
NARX (SimpleRNN)	86.65	115.68	90.39	116.87
LSTM Based NARX	133.41	168.11	117.48	148.66
Prophet	82.93	111.95	85.88	112.22
Prophet-LSTM Hybrid Model	62.17	80.48	65.58	84.73

When the performances of the models are evaluated with MAE and RMSE metrics over training and test data, it is seen that the models can generalise the data and there is no risk of overfitting in the models, except for the LSTM-based NARX model.

Findings & Feature Work

Key Findings:

- Synthetic datasets could failed to replicate time series data.
- When the performance metrics, loss function graphs are evaluated and the actual predicted consumption values are compared, it is seen that the Prophet-LSTM hybrid model is more successful in the prediction of hourly electricity consumption.

Future Work:

- Optimizing performance metrics
- Evaluation of short-term and long-term forecast performances
- Improving the performance of the model by selecting shorter and more significant periods.
- Updating the model regularly with new data and adapting to changing conditions.

THANK YOU!

Hayriye Yasak Ozkal

hayriye.yasakozkal@tcmb.gov.tr

Big Data Technologies Directorate

Central Bank of the Republic of Turkey

TÜRKİYE CUMHURİYET MERKEZ BANKASI İTÜ